CENTRE of POLICY STUDIES and the IMPACT PROJECT

Eleventh Floor
Menzies Building
PO Box 11E, Monash University
Wellington Road
CLAYTON Vic 3800 AUSTRALIA

Telephone: (03) 9905 2398, (03) 9905 5112
from overseas: 61 3 9905 5112 or 61 3 9905 2398

Fax: (03) 9905 2426
61 3 9905 2426

e-mail impact@vaxc.cc.monash.edu.au

web site http://www.monash.edu.au/policy/

SOLVING NONLINEAR ECONOMIC MODELS ACCURATELY VIA A LINEAR REPRESENTATION

by

K.R. PEARSON

Centre of Policy Studies, Monash University

Preliminary Working Paper No. IP-55 July 1991
Revised June 2002

ISSN 1 031 9034 ISBN 0 642 101053 1

The Centre of Policy Studies (COPS) is a research centre at Monash University devoted to quantitative analysis of issues relevant to Australian economic policy.

Explanatory Note, June 2002

This version (June 2002) is a revised version of the original IP-55 which was published in July 1991.

In July 1991, Ken Pearson worked at the

Mathematics Department, La Trobe University

and the

Impact Research Centre, University of Melbourne

The revisions made in this June 2002 version are minor. A small number of typographical errors were corrected. The paging and positioning of diagrams are slightly different from the original July 1991 version.

I am grateful to Brian Davey (Mathematics, La Trobe University) for technical assistance in producing the PDF file for the June 2002 version.

Ken Pearson

Ken.Pearson@BusEco.monash.edu.au

ABSTRACT

This paper focuses on one way a linearized representation of a nonlinear economic model can be used to obtain arbitrarily accurate solutions to simulations. The key is a method for translating a simulation problem directly to a so-called initial value problem. Since many different methods for solving initial value problems are known and well understood, and since each one converts to an algorithm for solving simulation problems, this insight greatly expands the computational tool kit for conducting simulations. This paper contains a survey of the theoretical results guaranteeing convergence and forming the basis for extrapolations of two important methods for solving initial value problems. Theoretical considerations suggest that the faster rate of convergence of one of these methods (the modified midpoint method) is likely to cause it to dominate the other (Euler's method) in many situations faced by applied general equilibrium modellers. The other main points of the paper are:

(i) to emphasize that linearized (symbolic) representations of models lead naturally to efficient algorithms which can be used to compute solutions having any desired degree of precision; and

(ii) to suggest that such accurate methods (rather than Johansen's method) should be the default when solving models (especially applied general equilibrium models) represented in linearized form.

TABLE OF CONTENTS

SOLVING NONLINEAR ECONOMIC MODELS ACCURATELY VIA A LINEAR REPRESENTATION

by

K.R. Pearson*

1. Introduction

Linearized representations have been used for over 30 years as a means of obtaining solutions of nonlinear economic models. Most commonly these representations have been used in applied general equilibrium analysis to obtain the so-called Johansen solution which is only an approximation to the exact solution (see, for example, Johansen (1960, 1974) and Dixon *et al.* (1982)). But, as shown in section 35 of Dixon *et al.* (1982), arbitrarily accurate solutions can be obtained by repeated solution of the linear equations associated with a linearized representation. This method, called Euler's method, has been used to solve ORANI and Miniature ORANI as described in sections 8, 33, 35 and 47 of Dixon *et al.* (1982); see also Exercises 3.7 and 3.8 of Dixon *et al.* (forthcoming). Although Euler's method was known to be closely related to the method of the same name for solving what are usually called initial value problems (see Problem Set 3B in Dixon *et al.* (forthcoming)), in the absence of a tight link between the two classes of problems, a direct proof of the convergence of the Euler solution to the true solution was given for simulation problems in section 35 of Dixon *et al.* (1982).

Initial value problems have been studied in great detail for many years. Many different methods for solving them are known and well understood both computationally and theoretically. Most books on numerical analysis or scientific numerical computation devote a chapter or more to these problems (see, for example, Chapter 15 of Press *et al.* (1986) or Chapter 8 of Dahlquist and Björck (1974)).

*I am grateful to P. Dixon, J.M. Horridge, A.Jones, J. Kautsky, M.Malakellis, A.Powell, I.Robinson and P.Wilcoxen for stimulating discussions and/or comments about different aspects of the material in this paper.

This paper shows how simulation problems can be converted directly and easily to initial value problems of exactly the kind studied in numerical analysis. A survey is given of two of the well-known methods for obtaining arbitrarily accurate solutions to initial value problems, and it is shown how these translate to algorithms for solving simulation problems with any desired accuracy. Included is a precise statement of theoretical results which (under suitable hypotheses on the functions involved) guarantee convergence and are the theoretical basis for extrapolation procedures (including those in Dixon *et al.* (1982) and in Exercise 3.8 of Dixon *et al.* (forthcoming)).

This reworking of the theoretical underpinnings for computations of so-called 'large-change' solutions of linearized models enables the techniques and experience of the numerical analysis literature on initial value problems to be directly appropriated. Each algorithm for solving an initial value problem potentially can be used to calculate the result of a simulation. When these methods are applied to a simulation, it turns out that the main computational task is the repeated solution of a system of linear equations. The numerical equations to be solved are those obtained naturally from a (symbolic) linearized representation of the model.

The methods used to solve initial value problems can produce solutions of any desired accuracy. They thus provide a means of calculating simulation results to any required level of precision.[1] Of course, different algorithms for the initial value problem become different algorithms for simulation problems.

The remainder of this paper is structured as follows. A brief introduction to the application of Euler's method in the simulation context is given in section 2, where this method's relatively slow speed of convergence is noted. In section 3, notation is set up for describing the computation of simulation results from a non-linear economic model; the term **simulation problem** is then defined precisely with the help of this notation. There follows in section 4 a precise definition of an **initial value problem**, and the statement of a theorem which identifies conditions sufficient for such a problem to possess a unique solution. The conversion

[1] It is tempting to say that these methods can calculate **exact** results but, given rounding errors, this is almost never true of any numerical computation. However, subject to the usual proviso that calculations carried out on computers nearly always contain rounding errors, it is reasonable to think that solving a simulation problem as an initial value problem gives access to the exact solution.

of a simulation problem to an initial value problem is then described. Sections 5 and 6 respectively contain reviews of two well-known methods for solving initial value problems; namely, Euler's method and the modified midpoint method. The use of Richardson's extrapolation is discussed in each case with the aid of a simple numerical example. Linear representations of a non-linear economic model are set out in section 7, where the practicalities of applying each of the foregoing methods are assessed in some detail. In the eighth and final section a tentative agenda for future research is sketched.

2. Euler's Method: Some Preliminary Remarks

As translated into the simulation context by Dixon *et al.* (1982), Euler's method divides the changes in the exogenous variables of the model into N equal parts (where N is an integer $\geqslant 1$) and results in an estimate $\underset{\sim}{z}_1^{(N)}$ for the solution to the simulation problem. The solution can be calculated to any desired accuracy because

$$\lim_{N\to\infty} \underset{\sim}{z}_1^{(N)} = \underset{\sim}{z}_1 \tag{2.1}$$

where $\underset{\sim}{z}_1$ is the vector of exact solutions of the simulation problem. Another way of saying this is that, given any desired (scalar) tolerance $\varepsilon > 0$, there is a (finite) value of N such that[2]

$$\|\underset{\sim}{z}_1^{(N)} - \underset{\sim}{z}_1\| < \varepsilon.$$

For this value of N, each component of $\underset{\sim}{z}_1^{(N)}$ will be at most ε in error from the exact value.

Despite being able to give arbitrarily accurate solutions, Euler's method is perhaps the worst of the many methods known for solving initial value problems. Here "worst" refers to the fact that the convergence in (2.1) is slow compared to other alternatives. Adapting one of these better methods for initial value problems to simulation problems will give a method

[2] In this paper, the vector norm $\|\ \|$ referred to is that defined by

$$\|(r_1, r_2, \ldots, r_k)\| = \max_{1\leqslant i\leqslant k} |r_i|,$$

although results stated would need very little modification to apply to any of several other norms. (See, for example, Section 4.1 of Gear (1971) for information about norms.)

which converges more rapidly to the true result. Another way of saying this is that significantly less computation will be required to give the desired accuracy.

For Euler's method, convergence in (2.1) is linear in the sense that the error between $z_1^{(N)}$ and z_1 is inversely proportional to N, so that, for example, multiplying N by 10 only reduces the error to one–tenth. Section 6 contains a description of the modified midpoint method which, for a given value of N, requires approximately the same amount of computation as Euler's method, but which has the distinct advantage that the error is inversely proportional to N^2. The difference between these two becomes even more striking when extrapolations are used. As shown in sections 5 and 6, extrapolations based on 3 Euler solutions have errors inversely proportional to N^3 while those based on the modified midpoint method are inversely proportional to N^6: then multiplying N by 10 reduces the errors to one-millionth of their previous size.

3. Models and Simulations

The methods described here apply to those economic models which can be written as a finite number of equations[3]

$$g_i(y_1, \ldots, y_{n+m}) = 0 \qquad i = 1, \ldots, m$$

where $g_1, \ldots, g_m$ are functions of the $m + n$ variables $y_1, \ldots, y_{n+m}$. To solve this model, n of the variables must be specified exogenously leaving the other m endogenous. We use $x_1, \ldots, x_n$ to denote a set of exogenous variables and $z_1, \ldots, z_m$ the corresponding endogenous ones. Then the equations of the model can be written as

$$g_i(z_1, \ldots, z_m, x_1, \ldots, x_n) = 0 \qquad i = 1, \ldots, m. \tag{3.1}$$

It is convenient to introduce vectors

$$\underset{\sim}{z} = (z_1, \ldots, z_m)$$

$$\underset{\sim}{x} = (x_1, \ldots, x_n)$$

$$\underset{\sim}{g}(\underset{\sim}{z}, \underset{\sim}{x}) = (g_1(\underset{\sim}{z}, \underset{\sim}{x}), \ldots, g_m(\underset{\sim}{z}, \underset{\sim}{x})).$$

Here $\underset{\sim}{g}$ is a vector-valued function from $\mathbb{R}^{n \times m}$ to $\mathbb{R}^m$ and the equations of the model are

$$\underset{\sim}{g}(\underset{\sim}{z}, \underset{\sim}{x}) = \underset{\sim}{0}. \tag{3.2}$$

Because $x_1, \ldots, x_n$ is a set of exogenous variables, the equations $\underset{\sim}{g}(\underset{\sim}{z}, \underset{\sim}{x}) = \underset{\sim}{0}$ determine $\underset{\sim}{z}$ as a function of $\underset{\sim}{x}$, say $\underset{\sim}{z} = \underset{\sim}{f}(\underset{\sim}{x})$. This means

$$z_i = f_i(\underset{\sim}{x}) = f_i\ (x_1, \ldots, x_n) \qquad i = 1, \ldots m$$

where[4]

$$\underset{\sim}{f}(\underset{\sim}{x}) = (f_1(\underset{\sim}{x}), \ldots, f_m(\underset{\sim}{x})).$$

[3] Models whose underlying theory involves inequalities as well as equations are not covered unless the inequalities can be rewritten as equations.

[4] In practice, although the functions $g_1, \ldots g_m$ are known (in the sense that formulae for them are given), explicit formulae for $\underset{\sim}{f}$ or $f_1, \ldots, f_m$ are not known; there is merely a guarantee of their existence.

It is assumed below that each function f_i has continuous first partial derivatives at each point $\underset{\sim}{x}$ in some relevant domain and also that each function g_i has continuous first partial derivatives with respect to each of its $m+n$ variables. That is, it is assumed that

$$\frac{\partial g_i}{\partial z_j} \quad \text{and} \quad \frac{\partial g_i}{\partial x_k}$$

exist and are continuous functions of $\underset{\sim}{z}$ and $\underset{\sim}{x}$ at each point in some relevant domain.

By a **simulation problem** for such as model we mean that one solution of the model, say $\underset{\sim}{z} = \underset{\sim}{z}_0$ when $\underset{\sim}{x} = \underset{\sim}{a}$, is given and also given is another set of values for the exogenous variables, say $\underset{\sim}{x} = \underset{\sim}{b}$. The problem is to calculate the value $\underset{\sim}{z}_1$ of $\underset{\sim}{z}$ when $\underset{\sim}{x} = \underset{\sim}{b}$. That is,

$$\begin{aligned} &\text{given } \underset{\sim}{z} = \underset{\sim}{z}_0 \text{ when } \underset{\sim}{x} = \underset{\sim}{a} \text{ is a solution of} \\ &(3.1) \text{ or } (3.2), \text{ find the value } \underset{\sim}{z}_1 \text{ of } \underset{\sim}{z} \text{ when } \underset{\sim}{x} = \underset{\sim}{b}. \end{aligned} \tag{3.3}$$

In the notation used above,

$$\underset{\sim}{g}(\underset{\sim}{z}_0, \underset{\sim}{a}) = \underset{\sim}{0} \quad (\text{or } \underset{\sim}{f}(\underset{\sim}{a}) = \underset{\sim}{z}_0)$$

and the problem is to find $\underset{\sim}{z}_1$ such that

$$\underset{\sim}{g}(\underset{\sim}{z}_1, \underset{\sim}{b}) = \underset{\sim}{0} \quad (\text{ or } \underset{\sim}{f}(\underset{\sim}{b}) = \underset{\sim}{z}_1).$$

We let

$$\underset{\sim}{a} = (a_1, \ldots, a_n) \quad \text{and } \underset{\sim}{b} = (b_1, \ldots, b_n).$$

4. Converting a Simulation Problem to an initial value Problem

An **Initial Value Problem** is a problem of the following form. Consider a vector $\underset{\sim}{w} = (w_1, \ldots, w_m)$ depending on a scalar variable $u \in \mathbb{R}$ such that $\underset{\sim}{w}$ is a differentiable function of u.[5] Given is a function $\underset{\sim}{q}(\underset{\sim}{w}, u)$ of $\underset{\sim}{w}$ and u such that

$$\frac{d\underset{\sim}{w}}{du} = \underset{\sim}{q}(\underset{\sim}{w}, u)$$

for all $\underset{\sim}{w}$ and u in some suitable domain.[6] Also given is the value $\underset{\sim}{w}_0$ of $\underset{\sim}{w}$ when $u = u_0$ (the "initial values") and another value, say u_1, of u. The problem is to calculate the value $\underset{\sim}{w}_1$ of $\underset{\sim}{w}$ when $u = u_1$. That is,[7]

$$\left.\begin{array}{l} \text{given} \quad \underset{\sim}{w}_0 = \underset{\sim}{w}(u_0), \\ \qquad\quad \dfrac{d\underset{\sim}{w}}{du} = \underset{\sim}{q}(\underset{\sim}{w}, u) \quad \text{for all relevant } \underset{\sim}{w}, u, \\ \text{the problem is to calculate } \underset{\sim}{w}_1 = \underset{\sim}{w}(u_1). \end{array}\right\} . \tag{4.1}$$

The function $\underset{\sim}{q}(\underset{\sim}{w}, u)$ is said to satisfy a **Lipschitz condition** if there is a positive constant L such that

$$\|\underset{\sim}{q}(\underset{\sim}{w}_1, u) - \underset{\sim}{q}(\underset{\sim}{w}_2, u)\| \leqslant L\|\underset{\sim}{w}_1 - \underset{\sim}{w}_2\|$$

for all u between u_0 and u_1 and for all vectors $\underset{\sim}{w}_1, \underset{\sim}{w}_2$.

The following well-known result (see, for example, Theorem 3.1 of Henrici (1962)) gives conditions which guarantee that an initial value problem has a unique solution.

Theorem 4.1. Consider the initial value problem given in (4.1). If $\underset{\sim}{q}(\underset{\sim}{w}, u)$ is continuous (for all $\underset{\sim}{w}$ and all u between u_0 and u_1) and if $\underset{\sim}{q}(\underset{\sim}{w}, u)$ satisfies a Lipschitz condition then there is a unique continuously differentiable function $\underset{\sim}{w} : [u_0, u_1] \to \mathbb{R}^m$ satisfying $\underset{\sim}{w}(u_0) = \underset{\sim}{w}_0$ and

$$\frac{d\underset{\sim}{w}}{du} = \underset{\sim}{q}(\underset{\sim}{w}, u) \qquad \text{for} \qquad u_0 \leqslant u \leqslant u_1 . \qquad \blacksquare$$

Although the Lipschitz condition as stated above must hold for all $\underset{\sim}{w}$, in practice it is only needed on some finite region of $\underset{\sim}{w}$ values within which the solution can be shown to lie. If

[5] That is, each w_i is a differentiable function of u.

[6] $\frac{d\underset{\sim}{w}}{du}$ is the vector $\left(\frac{dw_1}{du}, \ldots, \frac{dw_m}{du}\right)$ and $\underset{\sim}{q}$ is a function from $\mathbb{R}^{m+1}$ into $\mathbb{R}^m$.

[7] A simple example (in which $m = 1$) is: given $\frac{dw}{du} = w$ and $w = 1$ when $u = 0$, find w when $u = 1$. The solution is $w = e$ (since $w = e^u$ in general).

$\underset{\sim}{q}(\underset{\sim}{w}, u)$ has continuous partial derivatives with respect to $\underset{\sim}{w}$ then $\underset{\sim}{q}$ always satisfies a Lipschitz condition in any bounded region since L can be taken as the associated matrix norm of the matrix K_{ij} where

$$K_{ij} = \max_{(\underset{\sim}{w},u)\in\text{region}} \left| \frac{\partial q_i}{\partial w_j}(\underset{\sim}{w}, u) \right|$$

(see, for example, section 4.2 of Gear (1971)). In particular, if, as is often the case, $\underset{\sim}{q}(\underset{\sim}{w}, u)$ is an analytic function, then the initial value problem (4.1) has a unique solution.

The purpose of this section is to show that any simulation problem for a model satisfying the hypotheses of the previous sections can be converted to (or re-expressed as) an initial value problem. We use the notation introduced in the previous section for the simulation problem.

The key idea is to introduce a new scalar variable $v \in \mathbb{R}$ and to consider the exogenous variables $\underset{\sim}{x}$ as a function of v given by

$$\underset{\sim}{x} = \underset{\sim}{a} + v(\underset{\sim}{b} - \underset{\sim}{a}).$$

Note that $\underset{\sim}{x} = \underset{\sim}{a}$ when $v = 0$, $\underset{\sim}{x} = \underset{\sim}{b}$ when $v = 1$ and that $\underset{\sim}{x}$ moves along the straight line (in n dimensions) joining $\underset{\sim}{a}$ and $\underset{\sim}{b}$ as v increases from 0 to 1.

Because $\underset{\sim}{x}$ is a function of v and $\underset{\sim}{z}$ is a function of $\underset{\sim}{x}$ via $\underset{\sim}{z} = \underset{\sim}{f}(\underset{\sim}{x})$ or $\underset{\sim}{g}(\underset{\sim}{z}, \underset{\sim}{x}) = \underset{\sim}{0}$, it follows that $\underset{\sim}{z}$ is also a function of v. Indeed,

$$\underset{\sim}{z} = \underset{\sim}{f}(\underset{\sim}{x}) \quad \text{where } \underset{\sim}{x} = \underset{\sim}{a} + v(\underset{\sim}{b} - \underset{\sim}{a})$$

or, alternatively,

$$z_i = f_i(\underset{\sim}{x})$$

where

$$x_k = a_k + v(b_k - a_k) \text{ for } k = 1, \ldots, n.$$

By assumption each function f_i has continuous first partial derivatives while clearly each x_j is a differentiable function of v. Thus, by the Chain Rule (see, for example, Theorem 1 in section 8.7 of Kreyszig (1979), each z_i is a differentiable function of v and

$$\frac{dz_i}{dv} = \sum_{j=1}^{n} \frac{\partial f_i}{\partial x_j}\frac{dx_j}{dv} = \sum_{j=1}^{n} \frac{\partial z_i}{\partial x_j}\frac{dx_j}{dv}.$$

However, because no formulae for the functions $f_1, \ldots, f_m$ are given (there is merely a guarantee of their existence), the formula above does not give an effective way of calculating $\frac{dz_i}{dv}$. But we can go back to the original equations of the model to get an effective way of calculating these derivatives. Recall from (3.1) that

$$g_i(z_1, \ldots, z_m, x_1, \ldots, x_n) = 0 \quad \text{for} \quad i = 1, \ldots, m.$$

By assumption $\frac{\partial g_i}{\partial z_j}$ and $\frac{\partial g_i}{\partial x_k}$ exist and are continuous functions of $\underset{\sim}{z}$ and $\underset{\sim}{x}$. Also we have seen that each z_j is a differentiable function of v (as is each x_k). Thus the Chain Rule can be used to differentiate the above equations with respect to v. This gives

$$\sum_{j=1}^{m} \frac{\partial g_i}{\partial z_j} \frac{dz_j}{dv} + \sum_{k=1}^{n} \frac{\partial g_i}{\partial x_k} \frac{dx_k}{dv} = 0 \quad \text{for } i = 1, \ldots, m. \tag{4.2}$$

But $x_k = a_k + v(b_k - a_k)$ so that

$$\frac{dx_k}{dv} = b_k - a_k \quad \text{for } k = 1, \ldots, n.$$

Thus, expressing (4.2) in matrix form gives

$$\begin{bmatrix} \frac{\partial g_1}{\partial z_1} & \cdots & \frac{\partial g_1}{\partial z_m} \\ \vdots & & \vdots \\ \frac{\partial g_m}{\partial z_1} & \cdots & \frac{\partial g_m}{\partial z_m} \end{bmatrix} \begin{bmatrix} \frac{dz_1}{dv} \\ \vdots \\ \frac{dz_m}{dv} \end{bmatrix} = - \begin{bmatrix} \frac{\partial g_1}{\partial x_1} & \cdots & \frac{\partial g_1}{\partial x_n} \\ & \vdots & \\ \frac{\partial g_m}{\partial x_1} & \cdots & \frac{\partial g_m}{\partial x_n} \end{bmatrix} \begin{bmatrix} b_1 - a_1 \\ \vdots \\ b_n - a_n \end{bmatrix}.$$

That is,

$$C(\underset{\sim}{z}, \underset{\sim}{x}) \frac{d\underset{\sim}{z}}{dv} = D(\underset{\sim}{z}, \underset{\sim}{x})$$

where $\underset{\sim}{x} = \underset{\sim}{a} + v(\underset{\sim}{b} - \underset{\sim}{a}), C(\underset{\sim}{z}, \underset{\sim}{x})$ is the $m \times m$ matrix whose (i, j) entry is $\frac{\partial g_i}{\partial x_j}$ evaluated at $(\underset{\sim}{z}, \underset{\sim}{x})$, $\dfrac{d\underset{\sim}{z}}{dv}$ is the $m \times 1$ vector whose jth entry is $\frac{dz_j}{dv}$ and $D(\underset{\sim}{z}, \underset{\sim}{x})$ is the $m \times 1$ vector obtained by mutiplying the $m \times n$ matrix whose (i, k) entry is $\frac{\partial g_i}{\partial x_k}$ evaluated at $(\underset{\sim}{z}, \underset{\sim}{x})$ by the $n \times 1$ vector whose kth entry is $a_k - b_k$. Now $\underset{\sim}{x} = \underset{\sim}{a} + v(\underset{\sim}{b} - \underset{\sim}{a})$ depends on v. If we let

$$A(\underset{\sim}{z}, v) = C(\underset{\sim}{z}, \underset{\sim}{a} + v(\underset{\sim}{b} - \underset{\sim}{a})),$$

$$B(\underset{\sim}{z}, v) = D(\underset{\sim}{z}, \underset{\sim}{a} + v(\underset{\sim}{b} - \underset{\sim}{a}))$$

then, for all relevant $v \in \mathbb{R}$,

$$A(\underset{\sim}{z}, v) \frac{d\underset{\sim}{z}}{dv} = B(\underset{\sim}{z}, v). \tag{4.3}$$

For a given $\underset{\sim}{z}$ and v, this is a system of m linear equations in the m unknowns $\frac{dz_1}{dv}, \ldots, \frac{dz_m}{dv}$.

In practice, it seems to be the case that the matrix $A(\underset{\sim}{z}, v)$ is invertible for the relevant values of $\underset{\sim}{z}$ and v occurring for a valid closure in a well-specified model.[8] In what follows we assume that $A(\underset{\sim}{z}, v)$ is invertible at all relevant points. Then, for any relevant $\underset{\sim}{z}$ and v, we can solve (4.3) and so calculate $\dfrac{d\underset{\sim}{z}}{dv}$ (in terms of $\underset{\sim}{z}$ and v).[9]

Thus, as foreshadowed, the simulation problem has been converted to an initial value problem. Here $\underset{\sim}{z}$ is a differentiable function of v. The initial value problem is as follows.

$$\left.\begin{array}{l} \text{Given } \ \underset{\sim}{z} = \underset{\sim}{z}_0 \ \text{ when } \ v = 0, \ \text{and} \\ \text{that } \dfrac{d\underset{\sim}{z}}{dv} \text{ can be calculated by solving (4.3).} \\ \text{Find } \ \underset{\sim}{z}_1, \ \text{the value of } \underset{\sim}{z} \text{ when } v = 1. \end{array}\right\} \tag{4.4}$$

[8] I do not know of a theoretical result which guarantees the invertibility of this matrix. This is a question that needs further investigation. (The only examples I am aware of where $A(\underset{\sim}{z}, v)$ is not invertible at some point are cases where $\underset{\sim}{g}(\underset{\sim}{z}, \underset{\sim}{x}) = \underset{\sim}{0}$ does not have a unique solution for $\underset{\sim}{z}$ in terms of $\underset{\sim}{x}$ in a neighbourhood. This is not unlike the example in equation (35.6) of Dixon *et al.* (1982).)

[9] As explained in section 35.4 of Dixon *et al* (1982), we do not know *ex ante* whether (4.3) can be solved but numerical calculations can leave us confident *ex post*. (Also, while there is a very clear theoretical distinction between invertible and non-invertible matrices, the distinction is blurred in practice because of rounding errors in any actual calculation.)

5. Euler's Method for Initial Value Problems

The best-known method for solving an initial value problem is called Euler's method. Consider the problem

$$\left.\begin{array}{c}\text{Given: } \underset{\sim}{w} = \underset{\sim}{w}_0 \text{ when } u = u_0, \text{ and} \\ \dfrac{d\underset{\sim}{w}}{du} = \underset{\sim}{q}(\underset{\sim}{w}, u) \text{ for all } \underset{\sim}{w}, u. \\ \text{Calculate } \underset{\sim}{w}_1, \text{ the value of } \underset{\sim}{w} \text{ when } u = u_1.\end{array}\right\} \tag{5.1}$$

To calculate the solution by Euler's method, first select an integer $N \geqslant 1$. The idea is to divide the interval from u_0 to u_1 into N equal steps. Let

$$h = \frac{u_1 - u_0}{N}$$

and consider u values

$$u^{(s)} = u_0 + sh \quad \text{for} \quad 0 \leqslant s \leqslant N.$$

Note that $u^{(0)} = u_0$ and $u^{(N)} = u_0 + Nh = u_1$. Let $\underset{\sim}{w}^{(0)} = \underset{\sim}{w}_0$.

Given an estimate $\underset{\sim}{w}^{(s)}$ for $\underset{\sim}{w}$ when $u = u^{(s)}$, the derivative of $\underset{\sim}{w}$ at this point is used to calculate $\underset{\sim}{w}^{(s+1)}$, an estimate for the value of $\underset{\sim}{w}$ when $u = u^{(s+1)}$, via the formula

$$\underset{\sim}{w}^{(s+1)} = \underset{\sim}{w}^{(s)} + h\underset{\sim}{q}(\underset{\sim}{w}^{(s)}, u^{(s)}) \qquad \text{for} \quad 0 \leqslant s \leqslant N-1. \tag{5.2}$$

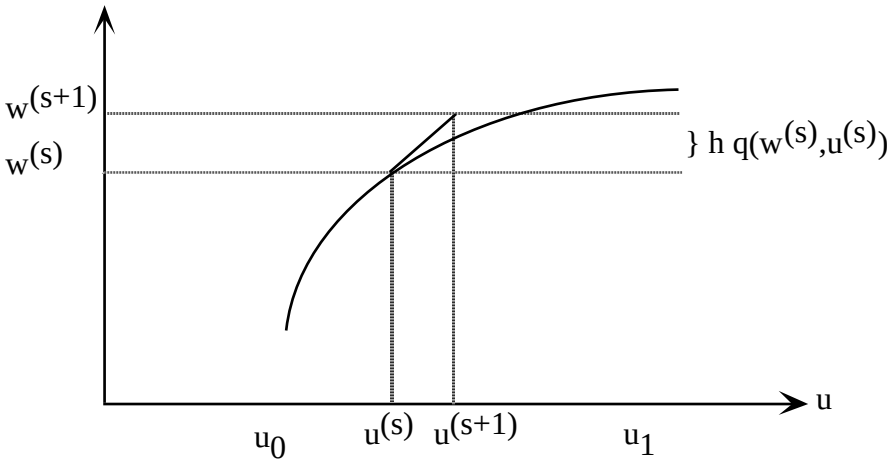
When $\underset{\sim}{w}$ is a scalar, this can be seen geometrically from the figure on the next page.

Although $\underset{\sim}{w}^{(1)}, \underset{\sim}{w}^{(2)}, \ldots, \underset{\sim}{w}^{(N)}$ are only estimates (or approximations) for the true values of $\underset{\sim}{w}$ when $u = u^{(1)}, u^{(2)}, \ldots, u^{(N)} = u_1$, and although (5.2) only gives an approximation for the true value of $\underset{\sim}{w}$ when $u = u^{(s+1)}$ even if $\underset{\sim}{w}^{(s)}$ is correct, it can be shown that, under suitable hypotheses on the function $\underset{\sim}{q}(\underset{\sim}{w}, u)$ (see Theorem 5.1 below), the value $\underset{\sim}{w}^{(N)}$ calculated as an estimate for $\underset{\sim}{w}_1$ approaches $\underset{\sim}{w}_1$ as $N \to \infty$. That is,

$$\lim_{N\to\infty} \underset{\sim}{w}^{(N)} = \underset{\sim}{w}_1.$$

Another way of saying this is that, given any desired tolerance $\varepsilon > 0$ (no matter how small), there is a (finite) value of N such that

$$\|\underset{\sim}{w}^{(N)} - \underset{\sim}{w}_1\| < \varepsilon.$$


w(s+1)
w(s)
} h q(w(s),u(s))
u
u0
u(s)
u(s+1)
u1

For this value of N, each component of $\underset{\sim}{w}^{(N)}$ will be at most ε in error from the exact value.

Clearly the estimate for $\underset{\sim}{w}_1$ actually calculated by Euler's method depends on N. Accordingly, let

$$\underset{\sim}{w}_1^E(N)$$

(where the E indicates Euler) denote the actual result obtained as an estimate for $\underset{\sim}{w}_1$ using Euler's method with N subintervals. The following theorem, which is a consequence of Theorem 3.1 of Gragg (1965), gives detailed information about $\underset{\sim}{w}_1^E(N)$.

Theorem 5.1. Consider a fixed initial value problem as in (5.1) above with true solution $\underset{\sim}{w}_1$. Suppose the function $\underset{\sim}{q}(\underset{\sim}{w}, u)$ is infinitely differentiable[10] and satisfies a Lipschitz condition. Then there are constant vectors $\underset{\sim}{c}_1, \underset{\sim}{c}_2, \ldots$ such that, for each integer $J \geqslant 1$,

$$\underset{\sim}{w}_1^E(N) = \underset{\sim}{w}_1 + \underset{\sim}{c}_1 \left(\frac{1}{N}\right) + \underset{\sim}{c}_2 \left(\frac{1}{N}\right)^2 + \cdots + \underset{\sim}{c}_J \left(\frac{1}{N}\right)^J + O\left(\frac{1}{N^{J+1}}\right). \qquad \blacksquare$$

As indicated below, this theorem guarantees convergence of $\underset{\sim}{w}_1^E(N)$ to $\underset{\sim}{w}$ (under the hypotheses stated) and is the theoretical basis for extrapolations. For convergence note that

$$\underset{\sim}{w}_1^E(N) = \underset{\sim}{w}_1 + O\left(\frac{1}{N}\right),$$

or, equivalently, there is a constant vector $\underset{\sim}{d}_1$ such that

$$\|\underset{\sim}{w}_1^E(N) - \underset{\sim}{w}_1\| \leqslant \frac{\|\underset{\sim}{d}_1\|}{N}.$$

This obviously guarantees that

$$\lim_{N\to\infty} \underset{\sim}{w}_1^E(N) = \underset{\sim}{w}_1.$$

In general, the error between what is calculated, $\underset{\sim}{w}_1^E(N)$, and the true result $\underset{\sim}{w}_1$ is given by the series

$$\underset{\sim}{c}_1 \left(\frac{1}{N}\right) + \underset{\sim}{c}_2 \left(\frac{1}{N}\right)^2 + \ldots \quad .$$

[10] In fact this hypothesis can be weakened. All that is required is tht the Jth Fréchet derivative exists and satisfies a Lipschitz condition; and note also that the Jth Fréchet derivative must exist if $\underset{\sim}{p}$ is J times continuously partially differentiable.

This is the theoretical basis for extrapolation, as is shown below. Because, in general, the vector $\underset{\sim}{c}_1$ in Theorem 5.1 is nonzero, the series there indicates that the rate of convergence of $\underset{\sim}{w}_1^E(N)$ to $\underset{\sim}{w}_1$ will only be inversely proportional to N. This means that the error in $\underset{\sim}{w}_1^E(2N)$ will be approximately half of the error in $\underset{\sim}{w}_1^E(N)$.

As an illustration, consider the function $w(u) = u^2$ with $u_0 = 1$ and $u_1 = 2$. Here

$$\frac{dw}{du} = 2u = 2\sqrt{w}. \tag{5.3}$$

Thus we can set up an initial value problem

$$\left.\begin{aligned} &u_0 = 1,\ w_0 = 1^2 = 1. \\ &\frac{dw}{du} = q(w,u). \\ &\text{Find}\ \ w_1 \text{ when } u_1 = 2. \end{aligned}\right\} \tag{5.4}$$

Equation (5.3) indicates two possible choices for the function q. It is possible to use either

$$q_1(w,u) = 2u \quad \text{or} \quad q_2(w,u) = 2\sqrt{w}.$$

Clearly the exact solution is $w_1 = 4$. In the table below are given the values of $w_1^E(N)$ for various values of N and for the two different functions q_1 and q_2.

N	1	2	10	20	100	1000
$w_1^E(N)$ using $q_1(w,u)$	3.0	3.5	3.9	3.95	3.99	3.999
$w_1^E(N)$ using $q_2(w,u)$	3.0	3.41421	3.86598	3.93185	3.98619	3.99862

It is easy to see that the errors are approximately inversely proportional to N. For example using $q_2(w,u)$, the error 0.13402 for $N = 10$ is reduced to 0.01381 (approximately divided by 10) when $N = 100$. (The fact that, for $q_1(w,u)$, the errors are exactly proportional to $1/N$ is not typical but a result of the very simple problem in this case.) The table makes it clear that the results obtained depend on the function $q(w,u)$ used. The results using q_2 are less accurate because calculations of $q_2(w^{(2)}, u^{(s)})$ in (5.2) depend on the $w^{(s)}$ values which are

only approximations to the true value of w when $u = u^{(s)}$ whereas $q_1(w^{(s)}, u^{(s)})$ only depends on $u^{(s)}$ which is known exactly.

Results such as the one in Theorem 5.1 occur frequently in numerical analysis and are the basis for various extrapolation procedures. Basically an extrapolation procedure is a method for combining results calculated for different values of N to produce a result which is significantly more accurate that any of the individual results. It can be thought of as a way of obtaining a new sequence of results whose convergence to the exact result is much faster than that of the original sequence. As indicated below, there are standard extrapolation procedures for obtaining cheaply from Euler results sequences which converge at rates inversely proportional to N^2, N^3, N^4 and so on. The extrapolations considered below are all called **Richardson** extrapolations.

The simplest Richardson extrapolation is based on two Euler solutions $\underset{\sim}{w}_1^E(N)$ and $\underset{\sim}{w}_1^E(2N)$. It follows from Theorem 5.1 that

$$\underset{\sim}{w}_1^E(N) = \underset{\sim}{w}_1 + \frac{\underset{\sim}{c}_1}{N} + O\left(\frac{1}{N^2}\right),$$
$$\underset{\sim}{w}_1^E(2N) = \underset{\sim}{w}_1 + \frac{\underset{\sim}{c}_1}{2N} + O\left(\frac{1}{N^2}\right),$$

and so

$$2\underset{\sim}{w}_1^E(2N) - \underset{\sim}{w}_1^E(N) = \underset{\sim}{w}_1 + O\left(\frac{1}{N^2}\right) \tag{5.5}$$

since the $\underset{\sim}{c}_1$ terms cancel out. Thus, if

$$\underset{\sim}{w}_1^{E2}(N) = 2\underset{\sim}{w}_1^E(2N) - \underset{\sim}{w}_1^E(N) \qquad \text{for all } N \geqslant 1, \tag{5.6}$$

the new sequence $\{\underset{\sim}{w}_1^{E2}(N)\}$ converges to $\underset{\sim}{w}_1$ at a rate inversely proportional to N^2. The values of $\underset{\sim}{w}_1^{E2}(N)$ are calculated by using Euler's method applied to N and $2N$ subintervals and then combining these results via (5.6). The superscript E2 is used to indicate that this is an extrapolation obtained from two different Euler solutions. We call it the **2-Euler-solution Richardson extrapolation**.

It is possible to obtain better extrapolations by combining three Euler approximations.

It follows from Theorem 5.1 that

$$\underset{\sim}{w}_1^E(N) = \underset{\sim}{w}_1 + \frac{\underset{\sim}{c}_1}{N} + \frac{\underset{\sim}{c}_2}{N^2} + O\left(\frac{1}{N^3}\right),$$
$$\underset{\sim}{w}_1^E(2N) = \underset{\sim}{w}_1 + \frac{\underset{\sim}{c}_1}{2N} + \frac{\underset{\sim}{c}_2}{4N^2} + O\left(\frac{1}{N^3}\right),$$
$$\underset{\sim}{w}_1^E(4N) = \underset{\sim}{w}_1 + \frac{\underset{\sim}{c}_1}{4N} + \frac{\underset{\sim}{c}_2}{16N^2} + O\left(\frac{1}{N^3}\right).$$

First choose pairwise combinations to eliminate the terms involving $\underset{\sim}{c}_1$. Thus

$$2\underset{\sim}{w}_1^E(2N) - \underset{\sim}{w}_1^E(N) = \underset{\sim}{w}_1 - \frac{\underset{\sim}{c}_2}{2N^2} + O\left(\frac{1}{N^3}\right),$$
$$4\underset{\sim}{w}_1^E(4N) - \underset{\sim}{w}_1^E(N) = 3\underset{\sim}{w}_1 - \frac{3\underset{\sim}{c}_2}{4N^2} + O\left(\frac{1}{N^3}\right).$$

Now the $\underset{\sim}{c}_2$ terms can be eliminated since

$$6[2\underset{\sim}{w}_1^E(2N) - \underset{\sim}{w}_1^E(N)] - 4[4\underset{\sim}{w}_1^E(4N) - \underset{\sim}{w}_1^E(N)] = -6\underset{\sim}{w}_1 + O\left(\frac{1}{N^3}\right).$$

That is,

$$-16\underset{\sim}{w}_1^E(4N) + 12\underset{\sim}{w}_1^E(2N) - 2\underset{\sim}{w}_1^E(N) = -6\underset{\sim}{w}_1 + O\left(\frac{1}{N^3}\right)$$

and so, dividing by -6,

$$\frac{8\underset{\sim}{w}_1^E(4N) - 6\underset{\sim}{w}_1^E(2N) + \underset{\sim}{w}_1^E(N)}{3} = \underset{\sim}{w}_1 + O\left(\frac{1}{N^3}\right).$$

So the 3-solution extrapolation is

$$\underset{\sim}{w}_1^{E3}(N) = \frac{8\underset{\sim}{w}_1^E(4N) - 6\underset{\sim}{w}_1^E(2N) + \underset{\sim}{w}_1^E(N)}{3}. \tag{5.7}$$

These values will converge to $\underset{\sim}{w}_1$ at a rate inversely proportional to N^3. Note that $\underset{\sim}{w}_1^{E3}(N)$ is the Richardson extrapolation based on the $N, 2N$ and $4N$ Euler results; it is called the **3-Euler-solution Richardson extrapolation**.

The table below shows extrapolation results for the example problem in (5.4), using $q_2(w,u) = 2\sqrt{w}$ (the less accurate of the two q's).

N	1	2	10	20
$w_1^{E2}(N)$	3.828427	3.948886	3.997712	3.999422
$w_1^{E3}(N)$	3.989039	3.998666	3.999992	3.999999

Note that here errors are approximately inversely proportional to N^2 for $w_1^{E2}(N)$ and to N^3 for $w_1^{E3}(N)$. For example, the error of 0.051114 in $w_1^{E2}(2)$ is reduced by a factor of approximately 100 to 0.000578 in $w_1^{E2}(20)$.

It is possible to form Richardson extrapolations based on 4 or more solutions (see, for example, section 7.2.2 of Dahlquist and Björck (1974)). For example, the 4-solution extrapolation is obtained by taking suitable combinations of the N, $2N$, $4N$ and $8N$ results to eliminate the $\underset{\sim}{c}_1, \underset{\sim}{c}_2$ and $\underset{\sim}{c}_3$ terms in these series. In this paper we concentrate on 2- and 3-solution extrapolations.

Some care must be exercised in using extrapolation results. Although $\underset{\sim}{w}_1^{E3}(N)$ will be considerably more accurate than $\underset{\sim}{w}_1^{E}(4N)$ for large N, it may not be a better approximation to $\underset{\sim}{w}_1$ for small N (say, $N = 1$). In a particular case it depends on whether most of the difference between $\underset{\sim}{w}_1$ and $\underset{\sim}{w}_1^{E}(N)$ is in the first one or two terms $\underset{\sim}{c}_1/N + \underset{\sim}{c}_2/N^2$ of the series or in the next. (Note that the individual entries in $\underset{\sim}{c}_2$ may be much larger in absolute value than the corresponding entries in $\underset{\sim}{c}_1$, and similarly for $\underset{\sim}{c}_3$ and so on.)

If indeed most of the error is in the $\frac{1}{N}$ term then

$$\underset{\sim}{w}_1^{E}(N) \doteqdot \underset{\sim}{w}_1 + \underset{\sim}{c}_1/N,$$
$$\underset{\sim}{w}_1^{E}(2N) \doteqdot \underset{\sim}{w}_1 + \underset{\sim}{c}_1/2N,$$
$$\underset{\sim}{w}_1^{E}(4N) \doteqdot \underset{\sim}{w}_1 + \underset{\sim}{c}_1/4N,$$

and so

$$\underset{\sim}{w}_1^{E}(2N) - \underset{\sim}{w}_1^{E}(4N) \doteqdot \underset{\sim}{c}_1/4N \doteqdot \frac{1}{2}(\underset{\sim}{w}_1^{E}(N) - \underset{\sim}{w}_1^{E}(2N))$$

or, taking the negative on both sides,

$$\underset{\sim}{w}_1^{E}(4N) - \underset{\sim}{w}_1^{E}(2N) \doteqdot \frac{1}{2}[\underset{\sim}{w}_1^{E}(2N) - \underset{\sim}{w}_1^{E}(N)]. \tag{5.8}$$

If, for example, one component of $\underset{\sim}{w}_1^{E}(2N)$ is larger than the corresponding component of $\underset{\sim}{w}_1^{E}(N)$ then this says that this component of $\underset{\sim}{w}_1^{E}(4N)$ should be still larger than this component of $\underset{\sim}{w}_1^{E}(2N)$ – indeed about half as large again. Similarly if one component of $\underset{\sim}{w}_1^{E}(2N)$

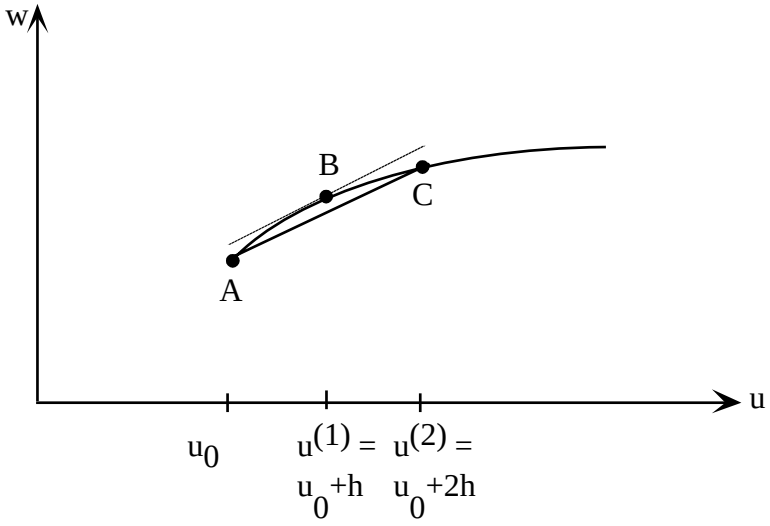
is less than the corresponding component of $\underset{\sim}{w}_1^E(N)$, then $\underset{\sim}{w}_1^E(4N)$ should be approximately half as much again smaller than $\underset{\sim}{w}_1^E(2N)$.

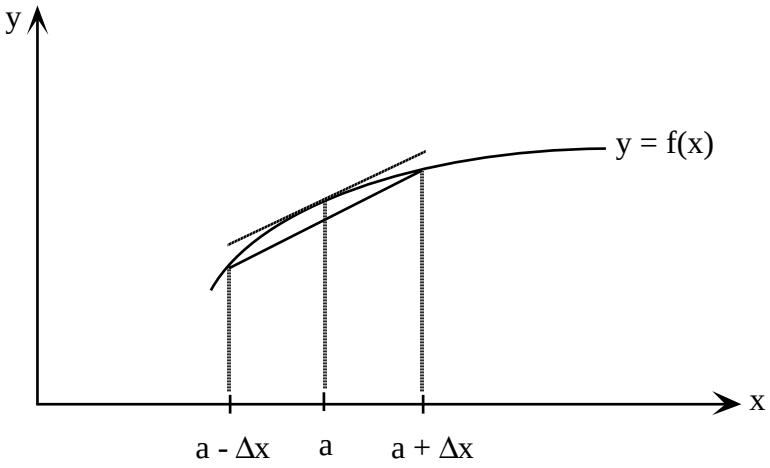
The following hypothetical examples show different possibilities.

	N-step	2N-step	4N-step	Comment
(1)	3.4612	3.5804	3.6485	(5.8) holds.
(2)	3.4612	3.5804	3.5926	4N result is bigger than 2N one, but increase is not approx half that from N to 2N.
(3)	3.4612	3.5804	3.5172	Oscillation. 2N larger than N but 4N less than N. This suggests 1/N term is not dominant for this N.
(4)	3.4612	3.3421	3.2802	(5.8) holds.
(5)	3.4612	3.3421	3.2015	4N result is less than 2N one, but decrease is not approx. half that from N to 2N.
(6)	3.4612	3.3421	3.3975	Similar to (3).
(7)	3.4612	3.5804	3.1214	Oscillation. But here difference between 4N and 2N significantly exceeds that between N and 2N. Perhaps results not converging?

(The results in (7) above could indicate that the values are not converging, perhaps because the hypotheses of Theorem 5.1 are not satisfied (or even because of an error in the code for computing the solutions). Alternatively, it may happen that (5.8) will be satisfied if N is increased significantly.)

When (5.8) holds, one can have **considerable confidence** that the 2-solution extrapolation is in error by no more than the absolute difference between these components of $\underset{\sim}{w}_1^E(N)$ and $\underset{\sim}{w}_1^E(2N)$ and can be **optimistic** (that is, have some confidence but not considerably confidence) that the 3-solution extrapolation result is a more accurate estimate of the exact solution than $\underset{\sim}{w}_1^E(4N)$. This is the basis of the "careful extrapolation" described in De Boor (1971). Of course, (5.8) may hold for some component of $\underset{\sim}{w}_1^E$ and not for others. A consequence of Theorem 5.1 is that (provided the hypotheses of that theorem are satisfied), (5.8) will eventually hold, provided N is large enough. So, if it does not hold for some value of N, the cautious approach is to increase N until it does hold and then base results on this larger value of N. Of course, extrapolated results may be (and often are) more accurate than the results on which they are based even if (5.8) does not hold.


w
B
C
A
u
u_0
$u^{(1)} = u_0 + h$
$u^{(2)} = u_0 + 2h$


y
y = f(x)
x
a - Δx
a
a + Δx

6. The Modified Midpoint Method for Initial Value Problems

Consider the initial value problem given in (5.1). As for Euler's method, the interval $[u_0, u_1]$ of interest is divided into N equal subintervals. The **modified midpoint method**[11] uses $N+1$ calculations to produce its estimate $\underset{\sim}{w}_1^M(N)$ of the value $\underset{\sim}{w}_1$. As before, let $h = (u_1 - u_0)/N$ and let $u^{(s)} = u_0 + sh$ for $0 \leqslant s \leqslant N$. Then, to calculate $\underset{\sim}{w}_1^M(N)$, put $\underset{\sim}{w}^{(0)} = \underset{\sim}{w}_0$,

$$\left.\begin{aligned} \underset{\sim}{w}^{(1)} &= \underset{\sim}{w}^{(0)} + h\underset{\sim}{q}(\underset{\sim}{w}^{(0)}, u^{(0)}), \\ \underset{\sim}{w}^{(s+1)} &= \underset{\sim}{w}^{(s-1)} + 2h\underset{\sim}{q}(\underset{\sim}{w}^{(s)}, u^{(s)}) \quad \text{for } 1 \leqslant s \leqslant N-1, \\ \underset{\sim}{w}^{(N+1)} &= \frac{1}{2}[\underset{\sim}{w}^{(N)} + \underset{\sim}{w}^{(N-1)} + h\underset{\sim}{q}(\underset{\sim}{w}^{(N)}, u^{(N)})]. \end{aligned}\right\} \tag{6.1}$$

Then $\underset{\sim}{w}_1^M(N) = \underset{\sim}{w}^{(N+1)}$.

Notice that $\underset{\sim}{w}^{(1)}$ is the same as in Euler's method. But then $\underset{\sim}{w}^{(2)}$ is based on $\underset{\sim}{w}^{(0)}$ and the derivative of $\underset{\sim}{w}$ with respect to u at $u = u^{(1)}$, as estimated by calculating $\underset{\sim}{q}(\underset{\sim}{w}^{(1)}, u^{(1)})$. Then $\underset{\sim}{w}^{(2)}$ is a good estimate of the value of $\underset{\sim}{w}$ at $u = u^{(2)}$ since, as can be seen from the first diagram on the previous page, the gradient of the tangent to the curve at $u = u^{(1)}$ (point P in the diagram) is a very good approximation to the gradient of the chord from A (when $u = u_0$) to C (where $u = u^{(2)}$).

Indeed (see the second diagram on the previous page), the calculation of $\underset{\sim}{w}^{(2)}, \ldots, \underset{\sim}{w}^{(N)}$ in the midpoint method is based on the idea that, for small Δx,

$$\frac{f(a + \Delta x) - f(a - \Delta x)}{2\Delta x} \tag{6.2}$$

is a very good approximation to the derivative $f'(a)$ of a function f at $x = a$.

Euler's method, on the other hand, relies on

$$\frac{f(a + \Delta x) - f(a)}{\Delta x} \tag{6.3}$$

as an estimate of $f'(a)$. The modified midpoint method is generally more accurate than Euler's method because (6.2) is generally a better estimate of $f'(a)$ than is (6.3).[12]

[11] This method was introduced by Gragg — see the method denoted by $A(t;h)$ at the end of section 5 of Gragg (1965). This method is sometimes referred to as Gragg's method (see, for example, section 6.3.2 of Stetter (1973)).

[12] The error in (6.2) is proportional to $(\Delta x)^3$ while that in (6.3) is proportional to $(\Delta x)^2$. (See, for example, Chapter 6 of Atkinson (1989).) Hence (6.2) is usually more accurate if Δx is small.

Information about the relation between $\underset{\sim}{w}_1^M(N)$ and the true solution $\underset{\sim}{w}_1$ is given by the following theorem which is a consequence of Theorem 6.3.1 in Stetter (1973). (See also section 5 of Gragg (1965) where this result was first proved.)

Theorem 6.1. (Gragg) Consider a fixed initial value problem as in (5.1) above with true solution $\underset{\sim}{w}_1$. Suppose the function $\underset{\sim}{q}(\underset{\sim}{w}, u)$ is infinitely differentiable[13] and satisfies a Lipschitz condition. Then there are constant vectors $\underset{\sim}{d}_1, \underset{\sim}{d}_2, \ldots$ such that, for all even integers N, and for all integers $J \geqslant 1$,

$$\underset{\sim}{w}_1^M(N) = \underset{\sim}{w}_1 + \underset{\sim}{d}_1 \left(\frac{1}{N}\right)^2 + \underset{\sim}{d}_2 \left(\frac{1}{N}\right)^4 + \cdots + \underset{\sim}{d}_J \left(\frac{1}{N}\right)^{2J} + O\left(\frac{1}{N^{2J+1}}\right).$$

■

Here the error between what is calculated, $\underset{\sim}{w}_1^M(N)$, and the true value $\underset{\sim}{w}_1$, is

$$\underset{\sim}{d}_1 \left(\frac{1}{N}\right)^2 + \underset{\sim}{d}_2 \left(\frac{1}{N}\right)^4 + \underset{\sim}{d}_3 \left(\frac{1}{N}\right)^6 + \ldots \quad .$$

There is no $\frac{1}{N}$ term and the dominant term involves $\left(\frac{1}{N}\right)^2$, and so the error is inversely proportional to N^2. If, for example, N is multiplied by 10, the error will decrease by a factor of approximately 100 (compared to only 10 for Euler's method). The fact that the modified midpoint method is a $1/N^2$ method while Euler's is a $1/N$ method suggests that $\underset{\sim}{w}_1^M(N)$ will be more accurate than $\underset{\sim}{w}_1^E(N)$. While this is usually the case, it is not guaranteed. (If the $\underset{\sim}{d}$'s in Theorem 6.1 are larger than the $\underset{\sim}{c}$'s in Theorem 5.1 for a particular problem, this may counteract the decrease from $1/N$ to $1/N^2$ for some values of N.)

The table below shows $w_1^M(N)$ for various values of N for the example problems (5.4), using $q(w, u) = 2\sqrt{w}$.

N	1	2	10	20	100	1000
$w_1^M(N)$	3.732051	3.892532	3.995037	3.998717	3.999950	3.9999995

It is easy to see that the errors here are approximately inversely proportional to N^2. Notice also, for example, that $w_1^M(2)$ is considerably more accurate than $w_1^E(2)$.

[13] The comments made in relation to Theorem 5.1 also apply here.

The absence of a $\left(\frac{1}{N}\right)^3$ term in the series means that extrapolations become more accurate more quickly. For example, if N is even,

$$\underset{\sim}{w}_1^M(N) = \underset{\sim}{w}_1 + \frac{\underset{\sim}{d}_1}{N^2} + O\left(\frac{1}{N^4}\right),$$
$$\underset{\sim}{w}_1^M(2N) = \underset{\sim}{w}_1 + \frac{\underset{\sim}{d}_1}{(2N)^2} + O\left(\frac{1}{N^4}\right)$$

and so

$$4\underset{\sim}{w}_1^M(2N) - \underset{\sim}{w}_1^M(N) = 3\underset{\sim}{w}_1 + O\left(\frac{1}{N^4}\right).$$

Hence the **2-modified-midpoint-solution Richardson extrapolation**

$$\underset{\sim}{w}_1^{M2}(N) = \frac{4\underset{\sim}{w}_1^M(2N) - \underset{\sim}{w}_1^M(N)}{3}$$

has error inversely proportional to N^4. (Recall that the 2-Euler-solution extrapolation has error inversely proportional to only N^2.)

A 3-solution extrapolation is also possible and the errors decrease in proportion to $1/N^6$. To see this, note that, if N is even,

$$\underset{\sim}{w}^M(N) = \underset{\sim}{w}_1 + \frac{\underset{\sim}{d}_1}{N^2} + \frac{\underset{\sim}{d}_2}{N^4} + O\left(\frac{1}{N^6}\right),$$
$$\underset{\sim}{w}^M(2N) = \underset{\sim}{w}_1 + \frac{\underset{\sim}{d}_1}{4N^2} + \frac{\underset{\sim}{d}_2}{16N^4} + O\left(\frac{1}{N^6}\right),$$
$$\underset{\sim}{w}^M(4N) = \underset{\sim}{w}_1 + \frac{\underset{\sim}{d}_1}{16N^2} + \frac{\underset{\sim}{d}_2}{256N^4} + O\left(\frac{1}{N^6}\right)$$

so that

$$4\underset{\sim}{w}^M(2N) - \underset{\sim}{w}^M(N) = 3\underset{\sim}{w}_1 - \frac{3\underset{\sim}{d}_2}{4N^4} + O\left(\frac{1}{N^6}\right),$$
$$16\underset{\sim}{w}^M(4N) - \underset{\sim}{w}^M(N) = 15\underset{\sim}{w}_1 - \frac{15\underset{\sim}{d}_2}{16N^4} + O\left(\frac{1}{N^6}\right),$$

and therefore

$$4[16\underset{\sim}{w}^M(4N) - \underset{\sim}{w}^M(N)] - 5[4\underset{\sim}{w}^M(2N) - \underset{\sim}{w}^M(N)] = 45\underset{\sim}{w}_1 + O\left(\frac{1}{N^6}\right).$$

Hence

$$\underset{\sim}{w}^{M3}(N) = \frac{4[16\underset{\sim}{w}^M(4N) - \underset{\sim}{w}^M(N)] - 5[4\underset{\sim}{w}^M(2N) - \underset{\sim}{w}^M(N)]}{45}$$
$$= \frac{64\underset{\sim}{w}^M(4N) - 20\underset{\sim}{w}^M(2N) + \underset{\sim}{w}^M(N)}{45},$$

the **3-modified-midpoint-solution Richardson extrapolation**, has error inversely proportional to N^6. Thus, provided of course that N is sufficiently large so that the $1/N^2$ and $1/N^4$ terms contain most of the errors in $\underset{\sim}{w}_1^M(N), \underset{\sim}{w}_1^M(2N)$ and $\underset{\sim}{w}_1^M(4N)$, it is reasonable to expect $\underset{\sim}{w}_1^{M3}(N)$ to be a good approximation to $\underset{\sim}{w}_1$ and that errors in it will decrease inversely proportionally to N^6.

The table below shows some 2-solution and 3-solution extrapolations for the initial value problem in (5.4) with $q(w,u) = 2\sqrt{w}$.

N	1	2	10	20
$w_1^{M2}(N)$	3.946026	3.995951	3.999991	3.9999994
$w_1^{M3}(N)$	3.999280	3.999920	3.999999989	3.9999999998

It is easy to check that errors in $w_1^{M2}(N)$ and $w_1^{M3}(N)$ here are (very) approximately proportional to $1/N^4$ and $1/N^6$ respectively. (Note, however, that the case $N = 1$ is not covered by Theorem 6.1.)

As with extrapolations based on Euler's method, we can only have confidence in $\underset{\sim}{w}_1^{M3}(N)$ if the $N, 2N$ and $4N$ step results are related appropriately. If most of the errors are in the $1/N^2$ terms then

$$\begin{aligned} \underset{\sim}{w}_1^M(N) &\doteq \underset{\sim}{w}_1 + \frac{\underset{\sim}{d}_1}{N^2}, \\ \underset{\sim}{w}_1^M(2N) &\doteq \underset{\sim}{w}_1 + \frac{\underset{\sim}{d}_1}{4N^2}, \\ \underset{\sim}{w}_1^M(4N) &\doteq \underset{\sim}{w}_1 + \frac{\underset{\sim}{d}_1}{16N^2} \end{aligned}$$

so that

$$\underset{\sim}{w}_1^M(4N) - \underset{\sim}{w}_1^M(2N) \doteq \frac{3}{16N^2} \doteq \frac{1}{4}\left[\underset{\sim}{w}_1^M(2N) - \underset{\sim}{w}_1^M(N)\right]. \tag{6.4}$$

This gives a test to apply to see how much confidence to place in the extrapolated result. Ideally N should be increased until (6.4) holds before extrapolating.

7. Linear Representations of an Economic Model

A simulation problem is converted to an initial value problem essentially by solving the system (4.2) of linear equations for $\frac{dz_j}{dv}$ (in terms of $\underset{\sim}{z}$ and v). In applying Euler's method (see (5.2)), it is necessary to calculate $\dfrac{d\underset{\sim}{z}}{dv}$ and multiply it by $h = 1/N$. (Recall that $v_0 = 0$ and $v_1 = 1$ when a simulation problem is converted to an initial value problem.) In applying the modified midpoint method (see (6.1)), $\dfrac{d\underset{\sim}{z}}{dv}$ is calculated and multiplied either by $h = \frac{1}{N}$ or $2h$. In each case, the quantity we multiply by can be thought of as a small change Δv in v and we have to calculate $\dfrac{d\underset{\sim}{z}}{dv}.\Delta v$. This leads naturally to the change linear representation of the model, discussed in Section 7.1 below. An alternative linear representation, based in percentage changes or proportional changes, is described in section 7.2. A combination of these two representations is often best, as described in section 7.3.

7.1 Change Linear Representation of a Model

To obtain the change linear representation of the economic model, begin with (4.2), multiply each equation by Δv and then write

$$\frac{dz_j}{dv}.\Delta v \text{ as } \Delta z_j \quad \text{and} \quad \frac{dx_k}{dv}.\Delta v \text{ as } \Delta x_k.$$

Thus the **change linear representation** of the economic model described by (3.1) is

$$\sum_{j=1}^{m} \frac{\partial g_i}{\partial z_j}\Delta z_j + \sum_{k=1}^{n} \frac{\partial g_i}{\partial x_k}\Delta x_k = 0 \qquad (i = 1, \ldots, m). \tag{7.1.1}$$

For a given Δv, Δx_k can be computed from

$$\Delta x_k = \frac{dx_k}{dv}\Delta v = (b_k - a_k)\Delta v.$$

In (7.1.1), the symbolic expressions for $\frac{\partial g_i}{\partial z_j}$, and $\frac{\partial g_i}{dx_k}$ involve $\underset{\sim}{z}$ and $\underset{\sim}{x}$ but not v. However, given v, it is easy to calculate $\underset{\sim}{x}$ from $\underset{\sim}{x} = \underset{\sim}{a} + v(\underset{\sim}{b} - \underset{\sim}{a})$. Thus, for a given $\underset{\sim}{z}$ and v, the numerical version of (7.1.1) can be computed and (provided the resulting system is non-singular) these equations can be solved to find values for the unknowns Δz_j $(j = 1, \ldots, m)$. These values can

then be fed into either the formulae (5.2) for Euler's method or (6.1) for the modified midpoint method.[14]

The equations (7.1.1) have a natural interpretation in terms of the model. If $\frac{\partial g_i}{\partial z_j}$ and $\frac{\partial g_i}{\partial x_k}$ are evaluated at some solution $(\underset{\sim}{z}, \underset{\sim}{x})$ of the model then, for small changes Δx_k in the exogenous variables x_k, the Δz_j's are the corresponding small changes in the endogenous variables z_j. Note also that (7.1.1) can be obtained by totally differentiating the left hand side of each equation in (3.1) and replacing dz_j and dx_k by Δz_j and Δx_k respectively.[15]

The change representation (7.1.1) can be combined with either Euler's method or the modified midpoint method to give an algorithm for solving the simulation problem (3.3). The two algorithms are described below.

7.1.1 Euler's Method for a Simulation Problem, Based on the Change Linear Representation

Given the simulation problem (3.3), Euler's method for the equivalent initial value problem (4.4) leads directly to the following algorithm for solving the simulation problem.

[14] Many other methods for initial value problems also rely on calculations of $\frac{dz_j}{dv}.\Delta v$ and so the change version linear representation (7.1.1) would be just as useful for them.

[15] If we totally differentiate each side of

$$g_i(z_1, \ldots, z_m,\ x_1, \ldots, x_n) = 0,$$

we obtain

$$0 = dg_i = \sum_{j=1}^{m} \frac{\partial g_i}{\partial z_j} dz_j + \sum_{k=1}^{n} \frac{\partial g_i}{\partial x_k} dx_k.$$

Algorithm 7.1.1. Let $\underset{\sim}{z}^{(0)} = \underset{\sim}{z}_0$ and $\underset{\sim}{x}^{(0)} = \underset{\sim}{a}$. For $s = 0$ to $N-1$,

(i) calculate the partial derivatives

$$\frac{\partial g_i}{\partial z_j} \quad \text{and} \quad \frac{\partial g_i}{\partial x_k} \quad \text{when } \underset{\sim}{x} = \underset{\sim}{x}^{(s)}, \underset{\sim}{z} = \underset{\sim}{z}^{(s)};$$

(ii) solve (7.1.1) for $\Delta z_j \quad (j = 1, \ldots, m)$, given

$$\Delta x_k = (b_k - a_k)/N \quad \text{for } k = 1, \ldots, n;$$

(iii) put

$$\underset{\sim}{z}^{(s+1)} = \underset{\sim}{z}^{(s)} + \Delta \underset{\sim}{z}, \qquad \underset{\sim}{x}^{(s+1)} = \underset{\sim}{x}^{(s)} + \Delta \underset{\sim}{x}$$

(where $\Delta \underset{\sim}{z}$ is the vector with components $\Delta z_1, \ldots, \Delta z_m$ and $\Delta \underset{\sim}{x}$ is the vector with components $\Delta x_1, \ldots, \Delta x_n$).

Finally let $\underset{\sim}{z}_1^E(N) = \underset{\sim}{z}^{(N)}$; this is the Euler estimate of the simulation result $\underset{\sim}{z}_1$. ■

Provided the hypotheses of Theorem 5.1 are satisfied,

$$\lim_{N \to \infty} \underset{\sim}{z}_1^E(N) = \underset{\sim}{z}_1.$$

Hence, by taking N large enough, $\underset{\sim}{z}_1^E(N)$ can be made arbitrarily close to $\underset{\sim}{z}_1$. Richardson extrapolations based on 2 or more Euler solutions can also be used to obtain quicker convergence. The theory outlined in section 5 gives the properties of these extrapolations.

7.1.2 Modified Midpoint Method for a Simulation, Based on the Change Linear Representation

Given the simulation problem (3.3), the modified midpoint method for the equivalent initial value problem (4.4) leads directly to the following algorithm for solving the simulation problem.

Algorithm 7.1.2. Let $\underset{\sim}{z}^{(0)} = \underset{\sim}{z}_0$ and $\underset{\sim}{x}^{(0)} = \underset{\sim}{a}$. Calculate $\underset{\sim}{z}^{(1)}$ and $\underset{\sim}{x}^{(1)}$ as in Algorithm 7.1.1.

For $s = 1$ to N,

(i) calculate the partial derivatives $\frac{\partial g_i}{\partial z_j}$ and $\frac{\partial g_i}{\partial x_k}$ when $\underset{\sim}{x} = \underset{\sim}{x}^{(s)}, \underset{\sim}{z} = \underset{\sim}{z}^{(s)}$;

(ii) solve (7.1.1) for $\Delta z_j \quad (j = 1, \ldots, n)$, given $\Delta x_k = (b_k - a_k)/N$ for $k = 1, \ldots, n$;

(iii) if $1 \leqslant s \leqslant N - 1$, put

$$\underset{\sim}{z}^{(s+1)} = \underset{\sim}{z}^{(s-1)} + 2\Delta\underset{\sim}{z}, \qquad \underset{\sim}{x}^{(s+1)} = \underset{\sim}{x}^{(s)} + \Delta\underset{\sim}{x},$$

or, if $s = N$, put

$$\underset{\sim}{z}^{(N+1)} = \frac{1}{2}[\underset{\sim}{z}^{(N-1)} + \underset{\sim}{z}^{(N)} + \Delta\underset{\sim}{z}]$$

where $\Delta\underset{\sim}{z}$ and $\Delta\underset{\sim}{x}$ are as in (iii) of Algorithm 7.1.1.

Finally let $\underset{\sim}{z}_1^M(N) = \underset{\sim}{z}^{(N+1)}$; this is the modified midpoint estimate of the simulation result $\underset{\sim}{z}_1$. ■

Again, provided the hypotheses of Theorem 6.1 are satisfied,

$$\lim_{N\to\infty} \underset{\sim}{z}_1^M(N) = \underset{\sim}{z}_1,$$

and Richardson extrapolations (whose behaviour was discussed in section 6) can be used to obtain more accurate results.

7.1.3 Data Requirements for the Change Representation

The data requirements are those values needed to calculate the coefficients in the system of equations (7.1.1) to be solved. Since $\frac{\partial g_i}{\partial z_j}$ and $\frac{\partial g_i}{\partial x_k}$ must be evaluated at various points in terms of $\underset{\sim}{x}$ and $\underset{\sim}{z}$, the (levels) values of $\underset{\sim}{x}$ and $\underset{\sim}{z}$ are needed at all points. Step (iii) of the algorithms keeps track of these values. Thus, for each variable z_j or x_k in the model, it is necessary to know their values at the initial solution.

Hence the change representation (7.1.1) naturally leads to algorithms which produce estimates of the values of all these variables at the final solution. They can also easily produce estimates of the actual change in each endogenous variable (by just adding up the changes at each step).

7.2 Percentage (or Proportional) Change Linear Representation of a Model

Consider a fixed (small) change Δv in v. Then, for any variable w (exogenous or endogenous), $\Delta w = \frac{dw}{dv}.\Delta v$ represents the corresponding small change in w. Provided $w \neq 0$, the corresponding percentage change pw in w is given by

$$pw = \frac{\Delta w}{w}.100 = \frac{\frac{dw}{dv}.\Delta v.100}{w}.$$

Indeed Δw can be recovered from pw via

$$\Delta w = \frac{w.pw}{100}.$$

Thus the change representation (7.1.1) becomes

$$\sum_{j=1}^{m} \frac{\partial g_i}{\partial z_j}.z_j.\frac{pz_j}{100} + \sum_{k=1}^{n} \frac{\partial g_i}{\partial x_k}.x_k.\frac{px_k}{100} = 0 \quad \text{for} \quad i = 1, \ldots, m. \tag{7.2.1}$$

This is the **percentage change linear representation** of the economic model described in (3.1). The closely related **proportional change linear representation** is obtained from (7.2.1) by omitting the two divisions by 100 and by intepreting pw as the proportional change $\frac{\Delta w}{w}$ in w. In what follows we concentrate on the percentage change representation, but everything said applies equally well (with the obvious minor changes) to the proportional change representation.

Since $\Delta w = \frac{w.pw}{100}$, Δw can be obtained from pw and, provided $w \neq 0$, pw can be obtained from Δw via $pw = \frac{\Delta w.100}{w}$. Thus the change of notation

$$\Delta w \leftrightarrow \frac{w.pw}{100}$$

is merely a different way of writing (7.1.1). That is, solutions of (7.1.1) can be recovered from solutions of (8.1.1) and vice versa. The natural interpretation of Δz_j as the approximate change in z_j resulting from small changes $\Delta x_1, \ldots, \Delta x_n$ in the exogenous variables means that pz_j can be interpreted as the approximate percentage change in z_j resulting from small percentage changes $px_1, \ldots, px_n$ in the exogenous variables.

7.2.1 Data Requirements for the Percentage Change Representation

The percentage change representation has the advantage that its data requirements are less than those of the change representation described in Section 7.1.3 above.

This is because there are usually many prices P, quantities X and associated dollar values F related via

$$F = PX \quad \text{or} \quad F - PX = 0. \tag{7.2.2}$$

Here the relevant function in (3.2) takes the form $g(F, P, X) = F - PX$, so $\frac{\partial g}{\partial F} = 1$, and $\frac{\partial g}{\partial P} = -X$. Thus the change version of (7.2.2) is

$$1.\Delta F - X.\Delta P - P.\Delta X = 0$$

$$\text{or} \qquad \Delta F = X.\Delta P + P.\Delta X. \tag{7.2.3}$$

The percentage change version is

$$1.F.\frac{pF}{100} - X.P.\frac{pP}{100} - P.X.\frac{pX}{100} = 0.$$

This can be simplified by replacing each occurrence of PX by F and then multiplying both sides by $100/F$ to obtain successively

$$\frac{F}{100}.pF - \frac{F}{100}.pP - \frac{F}{100}.pX = 0,$$

$$pF - pP - pX = 0, \quad \text{or} \quad pF = pP + pX. \tag{7.2.4}$$

Note that the values of P and X are not required in (7.2.4) but they are needed for (7.2.3).

Indeed the values of P and X can be eliminated from the percentage change version of other equations involving P and X. For example, a market clearing condition for commodities may be expressed as

$$X = X_1 + X_2 + X_3.$$

The change version is

$$\Delta X = \Delta X_1 + \Delta X_2 + \Delta X_3,$$

and so the percentage change version is

$$X.pX = X_1.pX_1 + X_2.pX_2 + X_3.pX_3.$$

Now, if both sides of this are multiplied by a relevant price P, this becomes

$$PX.pX = (PX_1).pX_1 + (PX_2).pX_2 + (PX_3).pX_3$$

or

$$F.pX = F_1.pX_1 + F_2.pX_2 + F_3.pX_3. \tag{7.2.5}$$

Then, provided the values of F, F_1, F_2, F_3 at this price are stored on the data base,[16] the values of the X's (and of P) are not needed.

This is a major advantage for the percentage change representation in comparison with the change representation. Provided dollar values (often at different prices) are held on the data base, individual prices and quantities are not needed.

Equally importantly, since

$$pF = pP + pX \text{ when } F = PX,$$

it is never necessary to have variables pF explicitly in the percentage change representation. (Simply replace pF by $pP + pX$ any time pF occurs.) The resulting reduction in the number of variables in the system of numerical equations (based on 7.2.1) to be solved speeds up actual calculations based on the percentage change representation.

Accordingly, in practice the percentage change representation (7.2.1) is written in such a way that

$$\left.\begin{array}{l}\text{the values of dollar values } F \text{ are held on the data base,}\\ \text{but not of prices } P \text{ and quantities } X\end{array}\right\} \tag{7.2.6}$$

and

$$\left.\begin{array}{l}\text{variables } pP \text{ and } pX \text{ occur explicitly,}\\ \text{but not } pF.\end{array}\right\} \tag{7.2.7}$$

Following (7.2.6) approximately halves the data requirements (when compared to the change version) while (7.2.7) reduces by approximately one-third the number of unknowns in the numerical equations to be solved.

[16] The term "data base" is used here to describe the data requirements of the method in question — that is, the set of values needed to calculate the coefficients of the system (7.2.1) of linear equations to be solved.

7.2.2 Update Formulae and Reporting Results

If a variable W is increased by the small amount ΔW then

$$\text{new}W = \text{old}W + \Delta W.$$

This is the way data is "updated" after each step in following Euler's method or the modified midpoint method based on the change representation. For the percentage change representation this becomes

$$\begin{aligned}\text{new}W &= \text{old}W + \frac{(\text{old}W).pW}{100} \\ &= \text{old}W\left(1 + \frac{pW}{100}\right).\end{aligned}$$

If $F = PX$, note that, since $pF = pP + pX$, this becomes

$$\text{new}F = \text{old}F\left(1 + \frac{pP + pX}{100}\right).$$

With the percentage change representation, it is most natural to report percentage changes in z_j (from its initial value in $\underset{\sim}{z}_0$ to its final value in $\underset{\sim}{z}_1$). It is also natural and easy to calculate and report final levels values of all data in the data base. After each step of the algorithm, data is updated using the update formulae discussed above. Thus the natural output from a simulation based on the percentage change representation of a model is

(a) the percentage changes in all endogenous variables as a result of the shocks, and

(b) the updated data base – that is, the (estimated) data values corresponding to the exogenous variables $\underset{\sim}{x} = \underset{\sim}{b}$.

When using the percentage change representation it is also most natural to specify the shocks (that is, movement of exogenous variables $\underset{\sim}{x}$ from $\underset{\sim}{a}$ to $\underset{\sim}{b}$) as percentage changes

$$px_k = \frac{(b_k - a_k)}{a_k} \qquad \text{(provided } a_k \neq 0\text{)}.$$

7.2.3 Euler's Method Based on the Percentage Change Representation

Given the simulation problem (3.3) and a percentage change representation (7.2.1) for the model, as modified by use of (7.2.6) and (7.2.7), Euler's method leads to the following algorithm.

Algorithm 7.2.1. Let $\underset{\sim}{z}^{(0)} = \underset{\sim}{z}_0$ and $\underset{\sim}{x}^{(0)} = \underset{\sim}{a}$.

For $s = 0$ to $N-1$,

(i) calculate the coefficients of (7.2.1) (as modified using appropriate versions of (7.2.6) and (7.2.7)) at $\underset{\sim}{x} = \underset{\sim}{x}^{(s)}, \underset{\sim}{z} = \underset{\sim}{z}^{(s)}$;

(ii) solve (7.2.1) as modified for $pz_j \quad (j = 1, \ldots m)$ given

$$px_k = \frac{[(b_k - a_k)/N].100}{a_k + \frac{s}{N}(b_k - a_k)};$$

(iii) put $\underset{\sim}{x}^{(s+1)} = \underset{\sim}{a} + \frac{s+1}{N}(\underset{\sim}{b} - \underset{\sim}{a})$ and

$$(\underset{\sim}{z}^{(s+1)})_j = (\underset{\sim}{z}^{(s)})_j + \frac{pz_j.(\underset{\sim}{z}^{(s)})_j}{100} \qquad \text{if } 1 \leqslant j \leqslant m.$$

(Here $(\underset{\sim}{z}^{(s)})_j$ denotes the jth component of the vector $\underset{\sim}{z}^{(s)}$.)

Then $\underset{\sim}{z}^{(N)}$ is $\underset{\sim}{z}_1^E(N)$. ■

7.2.4 Modified Midpoint Method Based on the Percentage Change Representation

Given the simulation problem (3.3) and a percentage change representation (7.2.1) for the model, as modified by use of (7.2.6) and (7.2.7), the modified midpoint method leads to the following algorithm.

Algorithm 7.2.2. Let $\underset{\sim}{z}^{(0)} = \underset{\sim}{z}_0$ and $\underset{\sim}{x}^{(0)} = \underset{\sim}{a}$. Calculate $\underset{\sim}{z}^{(1)}$ and $\underset{\sim}{x}^{(1)}$ as in Algorithm 7.2.1.

For $s = 1$ to N,

(i) calculate the coefficients of (7.2.1) (as modified using the appropriate versions of (7.2.6) and (7.2.7)) at $\underset{\sim}{x} = \underset{\sim}{x}^{(s)}, \ \underset{\sim}{z} = \underset{\sim}{z}^{(s)}$;

(ii) solve (7.2.1) as modified for pz_j $(j = 1, \ldots, m)$ given

$$px_k = \frac{[(b_k - a_k)/N].100}{a_k + \frac{s}{N}(b_k - a_k)};$$

(iii) if $1 \leqslant s \leqslant N-1$, put $\underset{\sim}{x}^{(s+1)} = \underset{\sim}{a} + \frac{s+1}{N}(\underset{\sim}{b} - \underset{\sim}{a})$ and

$$(\underset{\sim}{z}^{(s+1)})_j = (\underset{\sim}{z}^{(s-1)})_j + \frac{2.pz_j.(\underset{\sim}{z}^{(s)})_j}{100} \qquad \text{if } \ 1 \leqslant j \leqslant m,$$

or, if $s = N$, put

$$(\underset{\sim}{z}^{(N+1)})_j = \frac{1}{2}\left[(\underset{\sim}{z}^{(N-1)})_j + (\underset{\sim}{z}^{(N)})_j + \frac{pz_j.(\underset{\sim}{z}^{(N)})_j}{100}\right] \qquad \text{if } \ 1 \leqslant j \leqslant m.$$

Then $\underset{\sim}{z}^{(N+1)}$ is $\underset{\sim}{z}_1^M(N)$. ■

7.3 Mixed Percentage Change/Change Representation

Introduction of a percentage change variable pW is inappropriate if W can be zero. Thus quantities such as the balance of trade or tax rates (which can be positive, negative or zero) should never be replaced by their percentage changes; rather their change form ΔW should be left in the linear representation. For most other variables it is usually best (for the reasons outlined in section 7.2.1) to include their percentage change (and not their actual change) in the linear representation used. Thus, in practice, it is best to use a mixed representation which includes the percentage change pW of most variables and the change ΔW of a few variables.

Consider, for example, the levels equation

$$B = E - M \tag{7.3.1}$$

where E and M are respectively the foreign currency values of exports and imports and B is the trade balance. Because E and M are never zero but B can be zero, positive or negative, it is best to use pE, pM and ΔB as variables. The linearized representation of (7.3.1) is thus

$$\begin{aligned}\Delta B &= \Delta E - \Delta M \\ &= \frac{E.pE}{100} - \frac{M.pM}{100} = \frac{1}{100}(E.pE - M.pM).\end{aligned}$$

A second example involving a tax may help to tie together several points discussed in section 7.2 above, including (7.2.6) and (7.2.7). Consider a quantity X which has two relevant prices P_0, the basic price, and P_1, the price including an *ad valorem* tax levied at the fractional rate T, so that

$$P_1 = P_0(1 + T). \tag{7.3.2}$$

Following (7.2.6) the data base should contain $F_0 = P_0X$ and $F_1 = P_1X$ (not P_0, P_1, X or T). Following (7.2.7), and noting that ΔT rather than pT should appear as a variable, the variables that should appear explicitly in the linear representation are pX, pP_0 and ΔT. Including pP_1 explicitly is optional. If it is included, its connection to pP_0, pX and ΔT can be seen by totally differentiating (7.3.2); this gives

$$\Delta P_1 = (1 + T).\Delta P_0 + P_0.\Delta T$$

so that

$$P_1.pP_1 = (1+T).P_0.pP_0 + P_0.\Delta T.100$$

which, after multiplication of both sides by X, gives

$$F_1.pP_1 = F_1.pP_0 + 100F_0.\Delta T. \tag{7.3.3}$$

This is the appropriate linear representation of (7.3.2). However the variable pP_1 may not be of much interest and then it (and equation (7.3.3)) need not appear explicitly in the linear representation of the model. In this case, the update formula for F_1 is of some interest. The linear representation of

$$F_1 = F_0(1+T)$$

is

$$\begin{aligned}\Delta F_1 &= F_0.\Delta T + (1+T).\Delta F_0 \\ &= F_0.\Delta T + (1+T).\frac{F_0.pF_0}{100} \\ &= F_0.\Delta T + F_1\left(\frac{pP_0+pX}{100}\right).\end{aligned}$$

Thus

$$\begin{aligned}\text{new}F_1 &= \text{old}F_1 + F_0.\Delta T + \text{old}F_1\left(\frac{pP_0+pX}{100}\right) \\ &= \text{old}F_1\left(1+\frac{pP_0+pX}{100}\right) + F_0.\Delta T \end{aligned} \tag{7.3.4}$$

is the update formula for F_1 in terms of just the variables pP_0, pX and ΔT. (Of course if the variable pP_1 is also explicitly in the system, the update can also be accomplished via

$$\text{new}F_1 = \text{old}F_1\left(1+\frac{pP_1+pX}{100}\right)$$

which, from (7.3.3), is identical to (7.3.4).)

8. Topics for Further Investigation

This paper has set out the theory for converting a simulation problem to an initial value problem, and raised the possibility of applying different initial-value-problem methods (including, but not restricted to, Euler's method and the modified midpoint method) to simulation problems. This leaves many interesting and important questions unanswered, some of which are discussed briefly below.

1. *Achievable accuracy.* Although these methods can theoretically calculate solutions to any desired accuracy, in practice achievable accuracy is limited by machine precision and accumulation of rounding errors. What guidance can be given to determine the actual accuracy that can be achieved, and how many steps might be needed? (As the number of steps increases, results are theoretically more accurate. But, for a very large number of steps, rounding errors can more than offset this increase in accuracy – see, for example, section 1.3.4 of Gear (1971).)

2. *Which method?* The modified midpoint method is theoretically much more efficient than Euler's. In practice, what does this mean for applied general equilibrium models? Are there other methods for initial value problems which may give more efficient methods for solving simulations? (For the simulation problem, the cost of each evaluation of the function $\underset{\sim}{q}(\underset{\sim}{w}, u)$ in (4.1) is very high: this is because it requires calculation of the coefficients of a potentially large system of linear equations and then their solution. Accordingly, only methods with relatively small numbers of function evaluations are likely to be attractive.)

3. *Extrapolation methods.* There are several other possibilities besides the Richardson extrapolation described in this paper. Other possibilities are to use different meshes (rather than the $1, 2, 2^2, 2^3, \ldots$ mesh used by Richardson) or to use rational function extrapolation instead of the polynomials used in Richardson. (See, for example, Chapter 6 of Gear (1971) or Chapter 15 of Press *et al.* (1986).) Which is best for simulation problems?

4. *Are the methods robust?* That is, do the methods guarantee convergence in all or most circumstances (without user intervention)? Firstly this asks when the matrix $A(\underset{\sim}{z}, v)$ in (4.3) is invertible. This requires theoretical investigation (how is the invertibility of $A(\underset{\sim}{z}, v)$ related to the specification of a valid exogenous/endogenous split (or closure) for the model?) Secondly this asks when the hypotheses of Theorems 5.1 and 6.1, or a suitable modification of these,[17] hold. In particular, this requires theoretical investigation of the properties of $\frac{d\underset{\sim}{z}}{dv}$, as calculated by solving (4.3), as a function of $\underset{\sim}{z}$ and v. In some cases, or for some models and closures, it may be possible to give a theoretical guarantee that the conclusions of Theorem 5.1 and 6.1 apply for the simulation problem converted to an initial value problem. For some models, such a guarantee may not be available *a priori*; in such cases, section 35.4 of Dixon *et al* (1982) is highly relevant. Information on these questions can also be obtained "experimentally" by applying the methods to a wide class of models and closures. Already there is plenty of such experimental evidence that the invertibility of $A(\underset{\sim}{z}, v)$ is common (if not universal), and the preliminary information on convergence and extrapolation is positive. Thus there are reasonable grounds for hoping that these methods will prove very robust in practice.

5. *Stability.* Methods for initial value problems have different stability properties. How do these affect applications to simulation problems? For example, some methods are only weakly stable which means that small errors due to rounding or in early steps can introduce so-called parasitic solutions which can eventually dominate the true solution. (See, for example, Dahlquist and Björck (1974), Examples 8.3.1 and 8.3.8 or section 6.4 of Atkinson (1989).) This is one reason for preferring the modified midpoint method to the explicit (or unmodified) midpoint method (see, for example, section 8.3.1 of Dalquist and Björck (1974)).

6. *Variable step size.* In the discussion in sections 5 and 6, N is fixed in advance and the interval $[u_0, u_1]$ is divided into N equal subintervals. This amounts to taking a fixed step size equal to $(u_1 - u_0)/N$. However methods which vary the step size can achieve great

[17] As indicated earlier, the hypotheses of these theorems are sufficient conditions; they are not necessary.

efficiency by taking larger step sizes when the resulting errors are small and only taking small steps when larger ones would lead to unacceptably large errors (see, for example, Chapter 15 of Press *et al.* (1986)). Does varying step size have much to offer simulation applications?

One way to investigate these questions is to apply different methods to a range of applied general equilibrium models. The general-purpose economic modelling software GEMPACK (see Pearson (1988) and Codsi and Pearson (1988)) currently uses Euler's method with Richardson extrapolation. A prototype offering the modified midpoint method as an alternative has been used on several models. The preliminary results suggest that, in practice, this may prove considerably more efficient than Euler's method in most (if not all) cases. A striking example is the percentage change in US aggregate utility in the simulation described in Section V of Hertel *et al.* (1991): the value given by the 2-step modified midpoint method is more accurate than the 40-step Euler solution.[18]

[18] As discussed in Hertel *et al.* (1991), calculating this value is a good test for methods based on a linearized representation since the Johansen solution even has the sign incorrect. The correct value is -0.03621, the Johansen value is $+0.08280$, the 40-step Euler gives -0.03339 and the 2-step modified midpoint gives -0.03448.

REFERENCES

Atkinson, Kendall E. (1989), *An Introduction to Numerical Analysis*, 2nd edition, Wiley, New York.

Codsi, G. and K.R. Pearson (1988), GEMPACK: General-Purpose Software for Applied General Equilibrium and Other Economic Modellers, *Computer Science in Economics and Management*, Vol. 1, pp. 189-207.

Dahlquist, Germund and Ake Björck (1974), *Numerical Methods*, translated by Ned Andersen, Prentice-Hall, Englewood Cliffs.

de Boor, Carl (1971), CADRE: An Algorithm for Numerical Quadrature in: John R. Rice (ed.), *Mathematical Software*, Academic Press, New York, pp. 417-449.

Dixon, Peter B., B.R. Parmenter, John Sutton and D.P. Vincent (1982), *ORANI: A Multisectoral Model of the Australian Economy*, North-Holland, Amsterdam.

Dixon, P.B., B.R. Parmenter, Alan A. Powell and Peter J. Wilcoxen (forthcoming), *Notes and Problems in Applied Equilibrium Economics*, North-Holland, Amsterdam (publication expected in 1991).

Gear, C. William (1971), *Numerical Initial Value Problems in Ordinary Differential Equations*, Prentice-Hall, Englewood Cliffs.

Gragg, William B. (1965), On Extrapolation Algorithms for Ordinary Initial Value Problems, *J. SIAM Numerical Analysis*, vol. 2, pp. 384-403.

Henrici, Peter (1962), *Discrete Variable Methods in Ordinary Differential Equations*, Wiley, New York.

Hertel, Thomas, W., J. Mark Horridge and K.R. Pearson (1991), Mending the Family Tree: A Reconciliation of the Linearization and Levels Schools of CGE Modelling, *Impact Preliminary Working Paper*, No. IP-54, July, pp.45.

Johansen, Leif (1960, 1974), *A Multi-sectoral Study of Economic Growth*, first edition 1960, second edition 1974, North-Holland, Amsterdam.

Kreyszig, Erwin (1979), *Advanced Engineering Mathematics*, 4th edition, Wiley, New York.

Pearson, K.R. (1988), Automating the Computation of Solutions of Large Economic Models, *Economic Modelling*, vol. 5, pp. 385-395.

Press, W.H., B.P. Flannery, S.A. Tenkolsky and W.T. Vetterling (1986), *Numerical Recipes: The Art of Scientific Computing*, Cambridge University Press.

Stetter, Hans J. (1973), *Analysis of Discretization Methods for Ordinary Differential Equations*, Springer, Berlin 1973.